A Study on the Role of Data Mining In Enabling Efficient Decision Making In Organizations

Mohammad Shahid Kamal

Research Scholar (Computer Science), CMJ University, Shillong, Meghalaya

Abstract— In the subject matter of discussion the role of data mining and its utility in enabling efficient decision making in organizations is analysed in details. Firstly an introduction to data mining is given followed by the definition of the term in the traditional as well as the modern forms. Then the evolution of the different forms of data mining and its practical implications in organizations is also illustrated in details. In the overall context the advantages associated with data mining along with what the future holds is also explained in details. Finally a short and precise summary about data mining is also provided.

Index Terms— Data Mining, Decision Making.

---◆------------------------------------

1. INTRODUCTION TO MONTE CARLO METHODS

In the dictionary terms, mining means digging oil resources from the ground. From the context of data mining the data resources and the ground are mapped into data and knowledge respectively. It depicts a scenario where one uses data for useful knowledge. In the world of today it signifies the most common term related to data visualization and synchronization (Leondes, 2002).

Data mining is concerned with the discovery of patterns, associations, rules and other forms of knowledge in the data sets. This knowledge is extracted from the data rather than being formulated by the user (Sumathi and Sivanandam, 2006).

The progresses in digital data acquisition as well as storage technology has resulted in the growth of huge databases. This has occurred in all facets of human endeavour as well as celestial along with astronomical bodies. It typically deals with the data that has been collected rather than the purpose of data analysis. This precise leads to the point that the objectives of data mining exercise have no role in the data collection strategy. This is the difference where data mining differs from statistics in which data are collected from different strategies to answer specific questions.

Considering the other side of the coin when huge volumes of data are involved, lot of problems tend to arise also. Some of them relate on how to store the data as well as access the data in the first place. The relationships along with structures found within a set of data must be novel as well. It is often considered as the broader spectra of knowledge discovery in databases. In short data mining is an exercise in the interdisciplinary sense. All the components have a total impact on the final result of data mining (Hand, Mannila and Smyth, 2001).

In the sack of hay, data mining helps you to find needles. Data mining is the latest buzzword in the domain of database management and by the needle one refers to the single piece of information and the haystack is the large size of the data warehouse that is build over a period of time. In the world of today data mining products have captivated the industry by storm. In fact the major database vendors have already taken steps to make sure of the fact that the platforms incorporate the latest in data mining techniques.
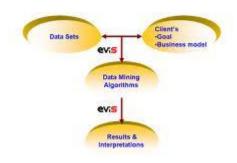
Figure 1: Flow of data mining from the organizational perspective

Through data mining one can discover the hidden treasure in the data warehouse which is a powerful technology with enormous amount of potential that helps the company to focus on important facets of information in their data warehouses. They tend to analyse futuristic behaviours and trends which allows the business to take proactive as well as knowledge based decisions. They can answer questions traditionally which took a long time to solve. The techniques can be implemented on the hardware as well as the software platforms to enhance the value of the existing sources of information. In addition they can be integrated with the new products as well as the systems and brought on a common level. The techniques involved in data mining evolve as a result of a long process of product development along with systematic research. In the business world of today it is finding more and more prominence as it is ably supported by the technologies.

WHAT IS DATA MINING?

Data mining happens to be one step in the data mining process. In fact the definition of data mining as well as the data mining process differs across various cultures. In a way data mining is the automated discovery of associations, clusters along with patterns in large sets of data and the exploitation of these discoveries to improve predictive modelling (Sumathi and Sivanandam, 2006).
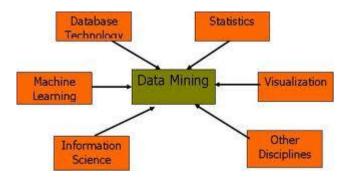

Figure 2: Process data warehouse to process mining

The data mining process consists of a series of steps which are repeated in an iterative fashion. Some of the process in this domain are data warehousing along with data preparation and cleaning. The cost involved in the data mining requirement is driven by the business benefit. It needs to be understood that since data mining is performed on a large set of data, it becomes quite difficult to apply it when algorithms whose execution increases as the data size grows. No wonder, lot of interest has been shown in tapping this data as it helps them extracting the information which may be of valuable use to the owner.

The discipline which precisely deals with the task is known as data mining. On the other side of the coin defining a scientific discipline is indeed a grave task and there is a point of dispute between the researches concerning the range as well as the limits of the study. In short data mining is the careful analysis of observational sets of data to determine the unsuspected levels of relation and then to summarize the data which makes it understandable as well as useful to the owner (Hand, Mannila and Smyth, 2001).

Earlier data mining is defined as traditionally the process of discovering patterns. This could be done automatically or semi automatically in large quantities of data and alongside the patterns need to be useful also. From the operational context data mining could be referred to as change which helps in enhancing future performance. Data mining is more of a practical term and does not imply theoretical learning in any way. In fact one is more interested in describing and finding structural patterns of data which will help them explain the data and draw useful information from it. Serious implications of data mining involves hundreds as well as thousands of cases on an individual basis. No doubt corporate data is a valuable asset and its utility has increased all the more with the enhancement as well as development of data mining techniques (Witten and Frank, 2005).

From the organizational perspective they provide a competitive advantage by the domain of data mining. Data mining is more concerned with predictive rather than retrospective models. It is not restricted to any specific industry and it requires the willingness and the use of intelligent technology to explore the correct knowledge in the embedded data. Data mining is also referred to as knowledge discovery in databases. It automatically involves searching large volumes of data for the benefit of association rules. Though it is a recent topic in the field of computer science but it applies many older computational techniques in the form of machine recognition as well as pattern learning. To look at it in another angle data mining is a relative new term, but the technology is not (Kirch, 2008)

DATA MINING AND EFFICIENT DECISION MAKING

Most of the organizations engage in data mining because of the following reasons

• To discover knowledge- The role of knowledge discovery is to determine explicit hidden relationships, patterns as well as correlations from the data stored in an enterprise's database

- Visualize data- The analysts need to make sense to the data which they are working upon. Before one gets to the basis of analyzing data they need to humanize the mass of data they are dealing and find a correct way to display the data as well.

- Correct data- quiet often it is found that organizations which deal with a large amount of data tend to have data at their disposal which is not complete as well as in contrast to the relevant information. The use of data mining techniques helps one to correct as well as identify information in the best possible way.

Effective data mining can improve the decision making in the organizations. All of us live in an information age where it is power. Data mining is also referred to as KDD which is definitely an integrative process. Once the information which is discovered is presented to the user, it calls for an evaluation where the mining process can be further refined. On the other hand data mining could also be called as knowledge mining in a lot of ways.

From the point of view of the organization the analysis of information is done at the top level of management. In the present world of today most of the organizations have operational systems and the difficult part is that no competitive advantage of any sort is derived from them in any way. Most of the companies have gathered a lot of information in the days gone by and have now realized the value of this hidden treasure. Whole hearted efforts are being made to channelize this information to improve the accuracy of the decision making process.

In the last decade or so the evolution of the world wide web has increased on a large scale. The improvement in the levels of communication along with the rise of the internet has contributed to this big time. The web world tends to improve the process as well as improves the quality of the decision to a considerable extent. In addition to this the effectiveness of the decision has improved to a considerable extent and this has paved way for the formation of data warehouses which supports the decision support systems. All this augers well in the days to come (Bhansali, 2010).

DIFFERENT TYPES OF DATA MINING

A host of analytic computer models have been used in data mining. The standard model in data mining techniques involves regression, decision trees along with neutral networks. These forms are not only the only methods but a host of methods have evolved in the recent past. Data mining requires the identification of a problem along with the collection of data which can pave way for better understanding. The tools employed in the case of data

mining needs to be accurate, predictable and capable of being implemented in the fastest possible time. In fact proper selection of data to the process of search is also very important. For this precise reason one needs to consider the point of data transformation also. If too many variables are involved it will lead to an information drain and too less an information will ignore important relationships in the variable (Olson and Delen, 2008).

Decision tree making is a common approach in data mining. When such a representation is adopted the various types of decisions could be implemented in a unified form. The edges of the decision making trees are represented by the fuzzy decision making models and the nodes are split according to the satisfaction of the formulas in the data records. They have considerable amount of impact in the field of machine learning as well as data mining. In fact decision making models tend to work better with a precise set of data, but it needs to be kept in mind that uncertainties have also a huge role in the data collection.

Figure 3: A typical example of a decision tree

As a data mining technique, link analysis has found widespread use. It concerns with the application of mathematical field in the concept of graphical theory. Some of the major strengths of this method involves it focuses on building relationships and is more useful from the visualization point of view. Further it creates derived statistics which can be used for further mining. In fact some data as well as data mining techniques involves the use of links as links are considered an effective tool to give shape to the data. It can pave way to new as well as useful form of data attributes (Linoff and Berry, 2011).

The types of data can be further grouped into three major data sets which are in the form of ordered data, record data as well as graph based data. As the data matures as well as develops, a greater variety of data is at the disposal which becomes available for disposal. In fact the difficult proposition in this regard is that all the major groupings are

not covered and other groupings are a definite possibility for sure (Tan, Kumar and Steinbach, 2006).

In the global age one tends to encounter different media types in multimedia data mining. They can be determined in terms of the dimensions of the space in which they find themselves in. They can be further segregated into further major media forms as follows

- dimensional data- this type of data is alphanumeric as well as regular in nature and a typical example in this regard would be the text data.

- Dimensional data- This data type has a one dimension of space imposed into the system. An example in this regard would be the audio data

- Dimensional data- It has two dimensions of space embedded on to it. Some of the common examples in this regard are the graphics as well as the imagery data

- 3 – Dimensional data- This type of data has three dimensions of space embedded on to it. Some of the common examples of this data are video as well as animation data.

The assimilation of data mining procedures into normal day to day activities has found common ground. It is concerned that the information is obtained at the cost of reduced privacy. In fact data mining techniques are used in fields like fraud detection in the world of today.

ADVANTAGES AND FUTURE OF DATA MINING

A host of advantages of data mining are as follows

- From the marketing perspective data mining aids direct marketers by providing accurate and correct information about the purchasing behaviour of the customers. This can help the marketers to direct their efforts towards the customers with precision. The retail stores can benefit from this mechanism in a lot of ways (Agarwal and Tayal, 2010).

- From the banking point of view data mining can assist financial institutions in the areas of loan information as well as credit reporting. For example the bank by comparing the attributes of a particular type of customers can associate the risk with each and every type of loan. It must be mentioned that though this system does not promise accuracy in all the cases, yes to a great extent it can help the credit reduce the losses.

- Legally data mining can aid the law enforcers by locating the criminals as well as apprehending them along with examining the location, size as well as the behavioural patterns of the criminals.

- From the point of view of the researches it can help them speed up their data analyze process. This in turn helps them to focus on more important projects

In fact data mining with decision trees has found a lot of prominence in the world of today. Decision trees tend to be self explanatory and are easy to follow when compacted. When compared to another level since the decision trees can be converted to a set of rules, the form of representation is considered all the more commendable. In addition to this the decision trees can handle both numerics as well as alpha data attributes and they can handle data sets that might have missing values also (Rokach and Maimon, 2008).

Since data mining projects in a majority of the cases are at the infancy stage, most of the data mining projects fall outside the preview of the boundaries of what the business managers expect (Rob and Coronel, 2009).

The ideal question doing the rounds is what does the future have in store for data mining. In the coming days much of the data mining may end up as standard tools build into the warehouse. On a parallel level the development of internet along with the information available has provided a competitive market for the products of various companies. In the world of today the availability of critical information is all the key point of consideration. In fact the magnitude of information available on the internet itself can improve the quality of the information. For any management to derive the optimum results it is imperative to focus on the information needs of the organization (Sumathi and Sivanandam, 2006).

The irony of the fact is that data mining is mainly used in deducting intrusions rather than to discover the knowledge as well as the nature of the attacks. In the coming days data mining can be used to improve IDS signatures as well as focus on the construction of alarm correlation systems also.

CONCLUSION

Traditionally speaking data mining requires the data to be housed in a single location. In fact recently established data mining systems finds it use in a wide variety of areas and also contributes to high performance networks. It is a well known technique for automatically and intelligent extracting information from plethora of data. In the midst of this it can provide personal information of the individuals which has a direct bearing on the privacy rights also. It can also reveal confidential business information about the

nature of business transactions compromising on the domain of free competition in the business setting.

The driving force happens to be one of the major policy changes in the present information age. The right to Privacy preserving data mining (PPDM) has been the latest buzzword in security as well as privacy research. From both the academic as well as industry domain a lot of interest has been generated. So it is quite clear that PPDM has evolved in a major form in the data management world. The fact of the matter is that the craze has evolved in the last decade or so and organizations have implemented it as a part of the curriculum which contributes to efficient decision making.

REFERENCES

1. Leondes, C. (2002). Database and Data Communication Network Systems: Techniques and Applications, Volume 1. USA: Elsevier.

2. Sumathi, S., and Sivanandam, S. (2006). Introduction to Data Mining and its Applications. Berlin: Springer.

3. Hand, D., Mannila, H., and Smyth, P. (2001). Principles of Data Mining. Massachusettes: MIT.

4. Witten, I., and Frank, E. (2005). Data Mining: Practical Machine Learning Tools and Techniques, Second Edition. London: Elsevier.

5. Kirch, W. (2008). Encyclopedia of Public Health: Volume 1: A - H Volume 2: I – Z. New York: Springer

6. Bhansali, N. (2010), Strategic Data Warehousing: Achieving Alignment with Business. NW: Taylor and Francis.

7. Olson, D., and Delen, D. (2008). Advanced Data Mining Techniques. Berlin: Springer.

8. Linoff, G., and Berry, M. (2011). Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management. IN: Wiley.

9. Tan, P., Kumar, V. and Steinbach, M. (2006). Introduction To Data Mining. New Delhi: Pearson.

10. Agarwal, B., and Tayal, S. (2009). Data Mining and Data Warehousing. New Delhi: Laxmi.

11. ROkach, L., and Maimon, Z. (2008). Data Mining with Decision Trees: Theroy and Applications. World Scientific.

12. Rob, P., Coronel, C. (2009). Database Systems: Design, Implementation, and Management. Massachusettes: Thomson.